TMA4267 - Linear Statistical Models 2009
Week 10: Test of Hypotheses and Confidence Intervals

Theoretical exercises

Rencher & Schaalje:

- 8.25

Computer exercises

Consider a model on the form

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + e_i, \quad i = 1, \ldots, n, \tag{1}$$

where $e_1, \ldots, e_n$ iid and $e_i \sim N(0, \sigma^2), \forall i$, y_i is the response and $\mathbf{x}_i$ covariates.

When gasoline is pumped into the tank of a car, vapors are vented into the atmosphere. An experiment was conducted to determine whether y, the amount of vapor, can be predicted using the following four variables based on initial conditions of the tank and the dispensed gasoline:

x_1:	`TankTemp`	tank temperature (^{o}F)
x_2:	`GasTemp`	gasoline temperature (^{o}F)
x_3:	`TankPres`	vapor pressure in tank (psi)
x_4:	`GasPres`	vapor pressure of gasoline (psi)

The data set `sniffer.dat` is available on the homepage. The exercise is based on problem 8.37 in Rencher & Schaalje (2008).

The hypothesis H_0 expresses the reduced model in terms of values of a subset of the β_j's in $\boldsymbol{\beta}$. The alternative hypothesis H_1 is associated with the full model.

a) Test the overall regression hypothesis $H_0 : \boldsymbol{\beta_1} = \mathbf{0}$ using (8.5) in Rencher & Schaalje (2008). The F-statistic can f.ex. be found by:

```
data = read.table("sniffer.txt",header=T)
model1 = lm(Y~.,data=data)
summary(model1)       # the output is explained in ?summary.lm
```

b) Test $H_0 : \beta_1 = \beta_3 = 0$, that is, that x_1 and x_3 do not significantly contribute above and beyond x_2 and x_4. The F-statistic can f.ex. be found by:

```
model2 = update(model1,.~.-TankTemp-TankPres)  # remove covariates
anova(model1,model2)     # the output is explained in ?avova.lm
```

c) Test $H_0 : \beta_j = 0$ for $j = 1,2,3,4$ using t_j in (8.40) in Rencher & Schaalje (2008). Use $t_{0.05/2}$ for each test. The t-statistics can f.ex. be found by:

```
summary(model1)
```

d) Find confidence intervals for $\beta_1, \beta_2, \beta_3$ and β_4 using both (8.47) in Rencher & Schaalje (2008).
R-command: `confint`.

e) Find the best set of predictors by removing the least significant predictors with `update` from the full model, one at a time, until all the predictors are significant at f.ex. $\alpha = 0.05$. Also try the `step`-function on the full model.

```
step(model1)
```

Compare the results.

The `step`-function minimize

$$\text{AIC} = n\log(\text{SSE}/n) + 2(p+1). \tag{2}$$

where n is the number of observations and p is the number of predictors. Lower AIC values indicate a better model.

f) A default argument in `step` is `direction="backward"`. Try

```
step(lm(Y~1,data=data),Y~TankTemp+GasTemp+TankPres+GasPres,
    direction="forward")
```

What is the difference compared to the `step`-procedure in e)?

When would it be favorable to use a forward-search and when would it be favorable to use a backward-search?

In e) we also used a backward-procedure when we used a significance level to remove predictors, could we also use a forward-procedure there?